Modern Day Evaluation of the Preston Curve:

The Relationship Between Life Expectancy and Income

Juan Garcia, Jasper Narvil, Soobin Oh

Econometric Analysis

Dr. Shatakshee Dhongde

Fall 2016

Abstract

The Preston Curve has long served as the foundation of global health policy with significant attention given to its implication that income has a causal effect on life expectancy. Our research sets to evaluate this relationship with modern data by incorporating econometric recommendations provided by scholars critical of the curve. We have added two primary extraneous variables into the relationship between life expectancy and income, health and education. Our models also test the role of unemployment and savings in this relationship, although it should be noted that these two specific factors are not the priority of this study. We test the reverse relationship as well, with income being the dependent variable rather than life expectancy. We hypothesize that these changes to the Preston Curve will update its validity and improve its accuracy.

I. Introduction

Through his iconic "Preston Curve," Preston suggests that income influences mortality, in which countries with higher national income per head tend to have higher life expectancies. This relationship flattens out throughout the years, and as it approaches the contemporary data. Written in 1975, Preston's study examines the relations by analyzing the data based on trends in 1900s, 1930s, and 1960s in which his argument held true. He suggests that the relationships might change as response to new influences, which sets basis for further research that prove or disprove the Preston Curve.

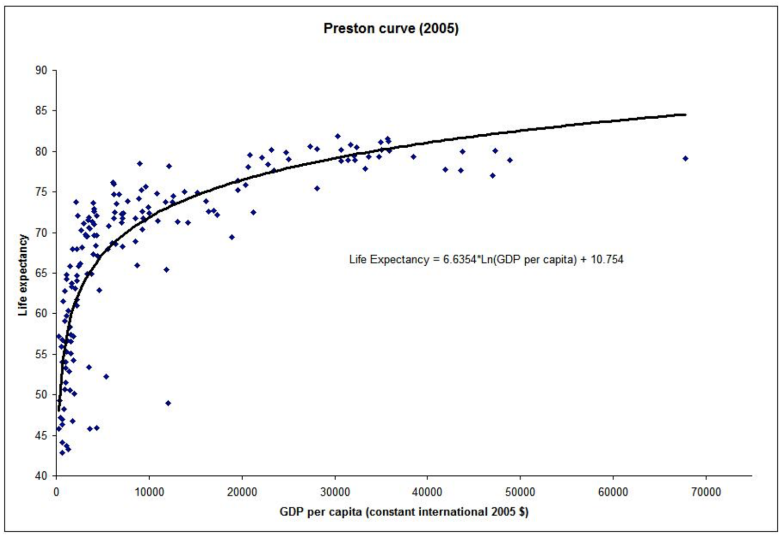

Figure 1: Preston Curve

Subsequent literature has sought to investigate these potential "new influences" to test their significance in influencing this famous curve. The Preston Curve has been a cornerstone of global health policy since Preston's findings were widely disseminated. Since then, neoliberal policy has been largely considered to be the harbinger of extended life expectancy through increasing per capita income. This assumption has been controversial with no consensus in the literature having yet been reached, however this policy has been widely accepted by legislators worldwide.

Updating policy to increase its efficacy in an increasingly globalized, changing world is crucial. We believe that the relationship between income and life expectancy is more nuanced than the simplified interpretations of the Preston Curve leaders which policy-makers operate under. We hypothesize that including factors such as education and health will produce a more accurate model of life expectancy as a function of income. We also expect that reversing the model to test the specific effects of life expectancy on income will produce strong, statistically significant results as well.

II. Literature Review

The complex relationship between income and life expectancy has been reviewed by a number of economic scholars to date. There is little consensus as to the principal, causal relationship between the two factors as exogenous entities such as health technology and education can be significantly correlated to both income and life expectancy.

Preston (1975) formally introduced the relationship between life expectancy and national income per had significantly shifted upward during the 20th century. He also argued that income growth influenced life expectancy for about 10-25%, and factors outside of income are more influential. Even for the influence attributed by level of income, he argues that the relationship occurs because of factors directly related to income level such as endogenous factors such as nutrition, medical and public health services, and literacy.

A paper by David Bloom and David Canning (2007) disseminates the key points on the amalgamated literature of interpretations of Preston's Curve. While several studies support his findings that income affects life expectancy to some degree, other scholars attribute these gains in life expectancy primarily to other factors that tend to also be highly correlated with income such as education and health technology. Other scholars saw Preston's correlation between life expectancy and income to have been interpreted incorrectly as a whole. Thomas and Schultz argue that health's effect on income creates the correlation and can more accurately be considered the driving causal factor, rather than the reverse. They argue that healthier workers are more productive, more likely to complete an education, and invest in their retirement and savings, all of which work to create a strong, robust economy.

Hussein et al (2014) look at life expectancy (as a measure of the health of a population) and its effect on per capita income and the growth rate of per capita income. They discover an inverse U-shaped relation, which suggests that increasing life expectancy eventually leads to a reduction in income. They argue that this is because increasing longevity increases the proportion of the elderly in the population, which can impose deadweight loss on the economy. They also find that life expectancy has a strong effect on per capita income independent of channels through which life expectancy is indirectly expected to operate (like savings, expenditure on health, etc.). This implies that there are other ways through which life expectancy operates to affect per capita income that have not yet been considered and that would provide important insight on growth policy. Importantly, the authors admit that longevity and health are not the same thing. Increasing longevity, according to their analysis, has negative consequences for economic growth. It is therefore important to evaluate relationships between income and other measures of health before concluding that investment in health does not sustain growth.

Our research will investigate the Preston curve and its criticisms by using modern data on per capita GDP, life expectancy, health, savings, unemployment, and education to test the dynamic relationships among these factors. We will conduct three multiple regressional analyses. Life expectancy is the dependent variable in the first regression as we test how it is affected by income, health, and tertiary education. The second uses this same model, but with the primary education enrollment variable instead of tertiary education enrollment. We are adding health into our regression as an explanatory variable instead of assuming it is represented by life expectancy. In doing this, we aim to test Hussein's assertion that life expectancy can act as a valid proxy to health.

Income is the independent variable first postulated by Preston to be a significant driver of life expectancy. To test the potential for reverse causality in this relationship, we use income as a dependent variable against life expectancy, life expectancy squared, savings, unemployment, and tertiary education in our final regression. The literature suggests this finding will be significant, so we hypothesize that the reverse causality model will have merits as well. Our findings will justify whether or not the multiple recommendations from previous literature more accurately explains the true relationship between income and life expectancy.

III. Data

This analysis conducts three regressions in order to evaluate whether income has a stronger effect on longevity or vice versa. We also incorporate several external factors to evaluate the roles of exogenous variables in the dynamic relationship between income and life expectancy. To examine this relationship, we regress life expectancy on health risk, log(per capita income), and gross tertiary education enrollment ratio. In a sister model, we regress life expectancy against health risk, log(per capita income), and primary education enrollment ratio. Literature on Preston's curve has established and implied strong relationships among the effects of income and education on life expectancy. Because primary education and tertiary education arguably do not have parallel effects on each societal dynamic we are incorporating, we test each individually to separately access their influence in the relationship.

As all of these variables are likely strong drivers of longevity, omitting them leads to biased estimators of the model parameters. Additionally, incorporating them allows us to explicitly control for these factors. We also use an index metric for health risk because "health" is very broad terminology. Longevity is not necessarily equivalent to health, as previous work on the reverse causality model of income on life expectancy seems to imply. Failing to discern between health and life expectancy and lead to inflated interpretations of the effect of income on life expectancy versus its effect on health. Additionally, we include gross savings as a percentage of GDP and unemployment to explain income. We believe both inclusion of both of these variables to be consistent with economic theory in determining income. As we believe expected longevity may influence an individual's decision to save for the future, omitting this variable may underspecify our model. Lastly, we use a log transformation on income in all of our models because our replication of the Preston Curve, represented in the figure below, illustrated a logarithmic relationship between income and life expectancy.

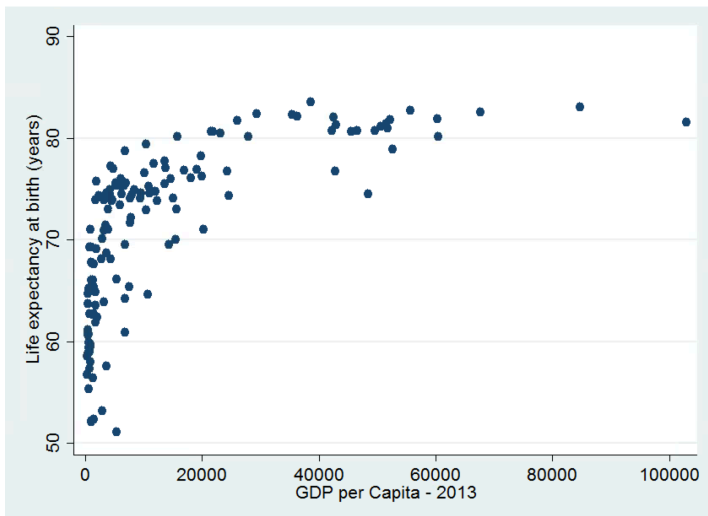

Figure 2. Our replication of the Preston Curve

A. Variable Descriptions

Variable Name	Description	Unit of Measurement	Type	Source
risk	Risk Penalty	None	Independent	United Nations, World Bank, World Health Organization
tertedu	Gross tertiary education enrollment ratio	Percentage of population	Independent	World Bank

primedu	Gross primary education enrollment ratio	Percentage	Independent	World Bank
le	Life expectancy at birth	Years	Independent /dependent	World Health Organization
le^2	Life expectancy at birth squared	Years squared	Independent	World Health Organization
savings	Gross savings as percentage of GDP	percentage	Independent	World Bank
lincome	Log of GDP per capita, 2013	US dollars/capita	Independent /Dependent	World Bank
unemployment	Percentage of population unemployed, 2013	percentage	Independent	World Bank

B. Summary Statistics

Variable	Observations	Mean	Std.Dev	Min	Max
risk	132	6.21947	0.9702383	3.07	8.95
tertedu	120	42.1686	26.25067	3.14233	95.34542
primedu	95	91.06885	10.65119	41.05925	99.94863
le	196	71.10032	8.582239	48.93793	83.83171
le^2	196	5128.534	1171.658	2394.921	7027.755
savings	150	22.00785	18.93167	-17.05493	195.6682
lincome	194	8.681153	1.487418	5.559237	11.64155
unemployment	172	8.572093	6.038718	0.3	31.1

C. Gauss-Markov Assumptions

We believe our model meets the five Gauss-Markov assumptions. The first assumption dictates that the model functions under linear parameters. We justify the relationship of each of our explanatory variables in our models on income and life expectancy through theoretical economics and prior literature on the topic. As we have not included any perfectly correlated variables, we believe this assumption holds in our three multiple regressions.

Assumption two states that the data points in our sample are random. We gathered a sample of the world's countries without any exclusionary conditions and consequently expect most of the data gathered to be free of inherent collection bias. Although there could be an inherent propensity of poorer countries to lack the resources to report a data point for every variable we used in comparison to wealthier countries, we found this effect to be minimal as most models used a relatively equitable amount of high and low income countries.

As can be seen in the correlation matrix below, none of our variables were perfectly correlated. The correlation matrix in the figure intuitively shows a negative correlation between risk and all other variables. Tertiary education shows a positive correlation with primary education, life expectancy, and log(income) as expected. The highest point of potential multicollinearity occurs between log(income) and and tertiary education. To test the effect of multicollinearity on the precision of our model, we conducted two F-tests to test the joint significance of Tertiary Education/Income and ___ . The results proved in both cases that the variables were jointly significant. The details of the tests can be found below in the "Robustness Tests" section. Thus, we assume to our data set to meet the third Gauss-Markov assumption.

```
. correlate risk tertedu primedu le le2 savings lincome unemployment
(obs=73)
```

	risk	tertedu	primedu	le	le2	savings	lincome	unempl~t
risk	1.0000							
tertedu	-0.0099	1.0000						
primedu	-0.3222	0.5606	1.0000					
le	-0.3525	0.7516	0.6739	1.0000				
le2	-0.3451	0.7585	0.6694	0.9976	1.0000			
savings	0.0172	-0.0311	0.1021	0.0591	0.0581	1.0000		
lincome	-0.1547	0.7820	0.5335	0.8498	0.8637	0.2213	1.0000	
unemployment	0.4051	0.0623	-0.1015	-0.0807	-0.0606	-0.0848	-0.0286	1.0000

Figure 3: Correlation Matrix

The fourth Gauss-Markov assumptions requires that the expected error value be zero, given any independent variable. The fifth assumption assumes property of homoskedasticity to be evident in our model, or that the variance is constant over all explanatory variables. Although both of these assumptions on our data and models cannot be proved directly, we believe our significant outputs and the economic rationale behind our methodology gives credence to the validity of our regression models in meeting these two assumptions.

IV. Results

A. Forward Simple Regression

$$\textbf{Life Expectancy} = \beta_0 + \beta_1\textbf{log(income)} + \textbf{u}$$

Our simple model defines life expectancy as a function of the log of GDP per capita (which we will refer to as 'income'). Calculating the regression coefficients yields the following function:

$$\textbf{le} = \textbf{30.42} + \textbf{4.68lincome} + \textbf{u}$$

The slope coefficient on log(income) indicates that a 1% increase in income is associated with an increase in life expectancy of 0.0468 years, or approximately 17 days. Our regression indicates that this coefficient is statistically significant but not practically (economically) significant. In other words, even though the rate of growth of income does have an effect on life expectancy, this effect does not appear to be large enough to warrant special importance. The reported R-Squared value of 0.6507 indicates that approximately 65% of the variation in life expectancy can be explained by the rate of growth of income, with the remaining 35% of the variation being explained by unaccounted factors.

B. Forward Multiple Regression

$$\textbf{Life Expectancy} = \beta_0 + \beta_1\textbf{log(income)} + \beta_2\textbf{Health Risk} + \beta_3\textbf{Tertiary Education} + \textbf{u}$$

This model tests the effects of income, health risk, and tertiary education on Life Expectancy. All coefficients were significant in this model at 10%, 5%, and 1% marks (See Appendix). The signs of the coefficients were as expected. Health risk showed a negative correlation to life expectancy while both tertiary education and income showed positive correlations to life expectancy. The R^2 value for this model was 0.7858 and significant overall.

$$\textbf{le} = \textbf{64.33} + \textbf{2.52lincome} - \textbf{3.18risk} + \textbf{0.126tertedu}$$

Our estimate thus shows that log(income) has a very small effect on life expectancy. The coefficient means that a 1% change in income leads to an estimated 0.0252 year increase in longevity (holding everything else constant), or approximately an additional 9 days. This is surprisingly practically insignificant: income does not have a large enough impact on life expectancy to be of real-world importance. The coefficient for health risk states that a 1 point increase in the index leads to a predicted decrease in life expectancy of 3.18 years, holding all other variables constant. The partial effect of an increase in 1 percentage point of tertiary education enrollment is an additional 0.126 years (approximately 46 days).

$$\textbf{Life Expectancy} = \beta_0 + \beta_1\textbf{log(income)} + \beta_2\textbf{Health Risk} + \beta_3\textbf{Primary Education} + u$$

Like the previous regression using tertiary enrollment as an explanatory variable, each coefficient in this model was significant at the 10%, 5%, and 1% levels with the overall model being significant at each level as well. The signs were consistent throughout this model and the previous one. However, the R^2-value increased from 0.7858 (Forward Regression A) to 0.8202 (Forward Regression B).

le = 36.54 + 3.78lincome -1.96risk+ 0.16primedu

Although the signs of the coefficient are consistent regardless of the type of enrollment being used as an explanatory variable, the size of the coefficients differ. The per-unit health risk penalty to life expectancy decreases by 1.22 years less using the forward regression model B versus the forward regression model A. In other words, as a country experiences a unit increase on the health risk scale under this model, life expectancy is predicted to decrease by less than it would if primary education is used as an explanatory versus using tertiary education. The coefficient for primary education itself is 0.16, an increase from tertiary's coefficient of 0.126 in the forward regression model A. In this model, a percentage point increase of primary education enrollment increases predicted life expectancy by 58 days (or 0.16 years). The log(income) coefficient states that a one percent increase in income leads to an estimated increase of 0.0378 years (14 days) in life expectancy. We can conclude that barring health risk, income, tertiary education, and primary education have statistically significant but practically insignificant effects on life expectancy. In other words, their partial effects are not large enough to generate real-world importance.

C. Reverse Simple Regression

$$\textbf{log(income)} = \beta_0 + \beta_1\textbf{Life Expectancy} + u$$

To examine the effect of life expectancy on income, we regressed the log of income with respect to life expectancy, receiving the following output:

lincome = -1.21+ 0.139le+ u

The intercept coefficient can be interpreted as the income when life expectancy is zero years. Thus, if the life expectancy at birth is zero years, income is e^-1.21 or approximately $0.30. However, this is not really a number of interest because life expectancy can't really be zero. The slope coefficient for life expectancy explains that for every 1 yr increase in longevity, there is a predicted increase of 13.9% in income. The corresponding t-statistic of 18.52 indicates that this coefficient is statistically significant, and the magnitude of the coefficient indicates that it has practical significance. That is, life expectancy does have an effect on the growth rate of income and this effect is large enough that it carries real-world importance. The R-Squared value of 0.6507 indicates that life expectancy explains roughly 65% of the variation in log of income, with the remaining 35% of the variation is due to unaccounted factors.

D. Reverse Multiple Regression

$$\textbf{log(income)} = \beta_0 + \beta_1\textbf{Life Expectancy} + \beta_2(\textbf{Life Expectancy})^2 + \beta_3\textbf{Unemployment} + \beta_4\textbf{Tertiary Education} + \beta_5\ \textbf{GrossSavings} + u$$

We are also interested in examining any partial effects of life expectancy on income. We hypothesize that increases in life expectancy does not exhibit constant percentage increases in income. As a result, we incorporate a squared term of life expectancy. We also include tertiary education enrollment as a percentage of the population, gross savings as a percentage of GDP, and the unemployment rate as determinants of income. We utilize these variables in order to reduce the risk of omitted variable bias and thus obtain an unbiased estimator for the effect of life expectancy on income.

Running the regression, we obtain the following functional form estimate:

lincome = 22.41 - 0.550 le + 0.005 le2 + -0.003 unemployment + 0.017 tertedu + 0.027 savings

The signs of the coefficients on savings and tertiary education enrollment align with our expectations. That is, we expect both to be positively correlated with income. The sign of unemployment also is expected, indicating that it has a negative effect on income. However, we encounter an outcome that our research did not predict when we examine the coefficients on life expectancy. Using calculus, we can determine the partial effect of life expectancy on income, holding the other variables constant. Mathematically,

$$\%\Delta\text{income} \approx 100(2\beta_2\ \text{LifeExpectancy} - \beta_1)\Delta\text{LifeExpectancy}$$

Or (using our coefficients)

$$\%\Delta\text{income} \approx 100[2(0.005)\ \text{LifeExpectancy} - 0.550]\Delta\text{LifeExpectancy}$$

$$= 100(0.01\text{xLifeExpectancy} - 0.551)\Delta\text{LifeExpectancy}$$

Thus, the relative change in income is dependent on life expectancy. This function says that income and life expectancy exhibit a U-shaped relation. At low values of life expectancy, the percent change in income is negative but is approaching zero. At high values of life expectancy, the relative change in income is positive and is quickly increasing. For example, at a life expectancy of 50 years, the predicted increase in income as a result of a one-year increase in longevity is -5.1%. When life expectancy is 70 years, the predicted change in income when longevity increases by one year is 14.9%. At what life expectancy does income growth become positive? Dividing $\boldsymbol{\beta_1/2\beta_2}$ produces 55, the life expectancy in years where income growth becomes positive. Perhaps unintuitively, unemployment proved to be non-significant in determining income.

Our estimates say that the effect of life expectancy on income is extremely practically significant, due to the calculations of the partial effect that we did earlier. In other words, we estimate that increases in life expectancy beyond 55 years of age lead to increasing returns to income, and at high levels of longevity the increases to income are very large holding all other factors constant. Because 88% of the countries in our data set have life expectancies greater than or equal to 55 years, we should not ignore the strong increasing effect of life expectancy. However, the sheer magnitude of the relative increases in income (in excess of 14% percent as shown above) very likely do not hold up in the real world, and it is most likely the case that our estimate has overestimated these effects. Husein et al found an upside down U-shaped relation between life expectancy and income, which is the opposite of what our estimate predicts.

The coefficient on savings indicates that a one percentage point increase in savings as a percentage of GDP leads to a predicted increase in income of 2.7%, holding other things constant. The coefficient on tertiary education enrollment ratio says that a one percentage point increase in that variable leads to a predicted increase in income of 1.7%, holding other things constant. These are practically significant because of their magnitude.

E. Robustness Test

Joint significance of explanatory variables can be examined by conducting the F-Test. In a multiple regression where life expectancy is the dependent variable, we tested the joint significance of log of income and the gross enrollment of tertiary education. Our unrestricted model is thus:

$$\textbf{Life Expectancy} = \beta_0 + \beta_1 \textbf{log(income)} + \beta_2 \textbf{Health Risk} + \beta_3 \textbf{Tertiary Education} + u$$

Our null hypothesis for the F-test is:

$$H_0 : \beta_1 = 0, \ \beta_3 = 0$$

The alternative hypothesis is that H_0 **is not true.**

In the unrestricted multiple regression of life expectancy on log(income), risk, and tertiary education, the R^2 value was 0.7858. When the tested variables log(income) and tertiary education enrollment were removed to create a restricted model, the R^2 value was 0.455. F-value resulted in 114.8 which is significantly greater than the critical value of 3.10 at 5%. We thus conclude that log(income) and tertiary education are jointly significant in the model estimate.

In a multiple regression where log of income is the dependent variable, we tested the joint significance of savings and unemployment. Our unrestricted model is:

$$\log(\text{income}) = \beta_0 + \beta_1\text{Life Expectancy} + \beta_2(\text{Life Expectancy})^2 + \beta_3\text{Unemployment} + \beta_4\text{Tertiary Education} + \beta_5\ \text{GrossSavings} + u$$

To compute the F-test statistic, we define our null hypothesis as follows:

$$H_0 : \beta_3 = 0,\ \ \beta_5 = 0$$

Which implies that the restricted model is:

$$\log(\text{income}) = \beta_0 + \beta_1\text{Life Expectancy} + \beta_2(\text{Life Expectancy})^2 + \beta_3\text{Tertiary Education} + u$$

Our alternative hypothesis is that H_0 **is not true.**

In calculating the f-value, we used the r-squared value for the unrestricted and restricted model for each of the regressions. We obtain an R^2 of 0.8185 for the unrestricted model, and an R^2 of 0.7517 for the restricted model. Thus, our F-test statistic is 17.86 which is statistically significant. Thus we conclude that Unemployment and Gross savings are jointly significant, even though unemployment by itself is statistically insignificant.

These variables were chosen to test the joint significance because they were highly correlated with each other (as in the case of log(inc) and tertiary education enrollment) or because one variable was strongly statistically significant while another was insignificant. Strong f-values would alleviate the concern for multicollinearity affecting the precision of our models, thus reinforcing the need for the variables in the respective models. In the case of the latter F-test, we can be assured that even though unemployment is statistically insignificant, we should probably still include it in our model.

V. Conclusion

Our results on the Forward Models representing the original Preston Curve indicate that all variables proved to be significant with high R^2 values. In adherence to the Preston Curve, the log of income had a small, positive coefficient and was significant at all levels. This finding both supports and negates the validity of the Preston Curve. Income, albeit with a positive relationship to life expectancy, influenced life expectancy to a much smaller degree than Preston assumed. Health risk and education proved to be wholly significant factors. This finding leads us to believe that these factors are just as important in determining life expectancy as income, if not more. Of the two forward models, we prefer model B to model A due to its higher R^2 value and conclude that primary education enrollment is a better determinant of a country's life expectancy than tertiary education. While most of the recommendations from the literature proved successful in creating a more valid model for the relationship between life expectancy and income, the fact that primary education model proved to be stronger than the tertiary model may imply that health technology is not the predominant predictor of life expectancy as the literature suggested. However, it should also be noted that determining causality in terms of education for this model would be difficult.

The reverse regression, however, produced more radical results. While education and savings had an intuitively positive relationship with income, life expectancy's effect was more complicated. Our regression output indicates that the coefficients are significant at all levels with the exception of the one corresponding to unemployment, which is surprisingly insignificant. More surprisingly, our results show that each added year of life expectancy directly correlated to decreasing income at a decreasing rate until age 55, at which it is zero. Beyond that age, it increases at an increasing rate. Because of this peculiar finding, we believe these results to be merely reflective of the correlation between low-income, short-life expectancy countries and high-income, high-life expectancy countries. These results may stem from the fact that our econometric tools failed to fully account for endogeneity, which occurs because income affects life expectancy which in turn affects income. Although most countries do have life expectancies above 55 at which income increases at an increasing rate, we conclude that our model for the reverse regression is not an accurate representation of the overall, global effect that life expectancy has on income.

While our results were nuanced, our hypothesis generally held true for the forward regression and is more than likely to be rejected in the reverse regression. The standing model for the Preston Curve should be updated with respect to education and risk to health so that policy-makers are made aware of the holistic effects on life expectancy outside of the small influence of income alone. For example, if a policy-maker decided to allow a factory to be built in an area to increase its GDP and subsequently life expectancy, the risks to health could negate those benefits to life expectancy contrary to the Preston Curve's consensus. Primary education should be valued in these decisions as well, as its role in determining life expectancy is arguably more prominent than that of higher education's. We did not find convincing proof that the reverse model testing life expectancy's effect on income is representative of causal relationships rather than correlative ones and do not hold the reverse causality to be true.

VI. References

- Bloom, D. E., & Canning, D. (2007). Commentary: The Preston Curve 30 years on: Still sparking fires. International Journal of Epidemiology, 36(3), 498-499. doi:10.1093/ije/dym079

- Husain, Z., Dutta, M., & Chowdhary, N. (2014). Is Health Wealth? Results of a Panel Data Analysis. Social Indicators Research, 117(1), 121-143. doi:http://dx.doi.org/10.1007/s11205-013-0337-4

- Preston, S. H. (1975). The Changing Relation between Mortality and Level of Economic Development. Population Studies, 29(2), 231. doi:10.2307/2173509

VII. Appendix

A. Estimation Results

Life Expectancy			
Independent Variables	Simple	Multiple A: tertedu	Multiple B: primedu
lincome	4.675376 (0.252509)	2.521184 (0.4624365)	3.777809 (0.293156)
risk		-3.180234 (0.4832963)	-1.960798 (0.4470114)
tertedu		0.1260612 (0.0252714)	
primedu			0.1620479 (0.0473976)
R^2	0.6507	0.7858	0.8208
n	186	95	94

Table 1. Note: the quantities in parentheses below the estimates are the standard errors

Log of Income		
Independent Variable	Simple	Multiple
le	0.1391849 (0.0075171)	-.5504452 (.1164314)
le2		.0047077 (.0008582)
tertedu		.0166577 (.0038814)
savings		.0265969 (.0055257)
unemployment		-.0031312 (.0097562)
R^2	0.6507	0.8185
n	186	103

Table 2. Note: the quantities in parentheses below the estimates are the standard errors

B. STATA outputs

1. Simple Regressions

```
. regress le lincome

      Source |       SS           df       MS      Number of obs   =       186
-------------+----------------------------------   F(1, 184)       =    342.83
       Model |  8973.49726         1  8973.49726   Prob > F        =    0.0000
    Residual |  4816.15055       184  26.1747312   R-squared       =    0.6507
-------------+----------------------------------   Adj R-squared   =    0.6488
       Total |  13789.6478       185  74.5386368   Root MSE        =    5.1161

------------------------------------------------------------------------------
          le |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lincome |   4.675376    .252509    18.52   0.000     4.177191    5.173561
       _cons |   30.42111   2.214115    13.74   0.000     26.05279    34.78943
------------------------------------------------------------------------------

. regress lincome le

      Source |       SS           df       MS      Number of obs   =       186
-------------+----------------------------------   F(1, 184)       =    342.83
       Model |  267.138918         1  267.138918   Prob > F        =    0.0000
    Residual |  143.375677       184  .779215635   R-squared       =    0.6507
-------------+----------------------------------   Adj R-squared   =    0.6488
       Total |  410.514594       185  2.21899781   Root MSE        =    .88273

------------------------------------------------------------------------------
     lincome |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          le |   .1391849   .0075171    18.52   0.000      .124354    .1540157
       _cons |  -1.215974   .5363155    -2.27   0.025    -2.274093   -.1578553
------------------------------------------------------------------------------
```

1. Multiple Regressions

```
. regress le lincome risk tertedu

      Source |       SS           df       MS      Number of obs   =        95
-------------+----------------------------------   F(3, 91)        =    111.31
       Model |  5392.22565         3  1797.40855   Prob > F        =    0.0000
    Residual |  1469.48107        91  16.1481436   R-squared       =    0.7858
-------------+----------------------------------   Adj R-squared   =    0.7788
       Total |  6861.70672        94    72.99688   Root MSE        =    4.0185

------------------------------------------------------------------------------
          le |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lincome |   2.521184   .4624365     5.45   0.000     1.602611    3.439757
        risk |  -3.180234   .4832963    -6.58   0.000    -4.140243   -2.220225
     tertedu |   .1260612   .0252714     4.99   0.000     .0758627    .1762596
       _cons |   64.33272   4.827325    13.33   0.000     54.74383    73.92161
------------------------------------------------------------------------------

. regress le lincome risk primedu

      Source |       SS           df       MS      Number of obs   =        94
-------------+----------------------------------   F(3, 90)        =    137.45
       Model |  4709.41436         3  1569.80479   Prob > F        =    0.0000
    Residual |  1027.87106        90  11.4207896   R-squared       =    0.8208
-------------+----------------------------------   Adj R-squared   =    0.8149
       Total |  5737.28542        93   61.691241   Root MSE        =    3.3795

------------------------------------------------------------------------------
          le |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lincome |   3.777809    .293156    12.89   0.000     3.195403    4.360215
        risk |  -1.960798   .4470114    -4.39   0.000    -2.848864   -1.072732
     primedu |   .1620479   .0473976     3.42   0.001     .0678843    .2562115
       _cons |   36.54366    5.21818     7.00   0.000     26.17684    46.91049
------------------------------------------------------------------------------
```

```
. regress lincome le le2 savings tertedu unemployment

      Source |       SS       df       MS              Number of obs =     103
-------------+------------------------------           F(5, 97)      =   87.48
       Model |  177.471052     5  35.4942103           Prob > F      =  0.0000
    Residual |  39.3547822    97  .405719404           R-squared     =  0.8185
-------------+------------------------------           Adj R-squared =  0.8091
       Total |  216.825834   102  2.12574347           Root MSE      =  .63696

------------------------------------------------------------------------------
     lincome |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          le |  -.5504452   .1164314    -4.73   0.000    -.7815293    -.319361
         le2 |   .0047077   .0008582     5.49   0.000     .0030045    .0064109
     savings |   .0265969   .0055257     4.81   0.000       .01563    .0375638
     tertedu |   .0166577   .0038814     4.29   0.000     .0089542    .0243613
unemployment |  -.0031312   .0097562    -0.32   0.749    -.0224945    .0162321
       _cons |   22.40876   3.963296     5.65   0.000     14.54272    30.27481
------------------------------------------------------------------------------
```

C. Raw data

Countries (A-K)				
Afghanistan	Benin	Chile	Equatorial Guinea	Guinea-Bissau
Albania	Bermuda	China	Eritrea	Guyana
Algeria	Bhutan	Colombia	Estonia	Haiti
American Samoa	Bolivia	Comoros	Ethiopia	Honduras
Andorra	Bosnia and Herzegovina	Congo, Dem. Rep.	Faroe Islands	Hong Kong SAR
Angola	Botswana	Congo, Rep.	Fiji	Hungary
Antigua and Barbuda	Brazil	Costa Rica	Finland	Iceland
Argentina	British Virgin Islands	Cote d'Ivoire	France	India
Armenia	Brunei Darussalam	Croatia	French Polynesia	Indonesia
Aruba	Bulgaria	Cuba	Gabon	Iran, Islamic Rep.
Australia	Burkina Faso	Curacao	Gambia, The	Iraq
Austria	Burundi	Cyprus	Georgia	Ireland
Azerbaijan	Cabo Verde	Czech Republic	Germany	Isle of Man
Bahamas, The	Cambodia	Denmark	Ghana	Israel
Bahrain	Cameroon	Djibouti	Gibraltar	Italy
Bangladesh	Canada	Dominica	Greenland	Jamaica
Barbados	Cayman Islands	Dominican Republic	Grenada	Japan
Belarus	Central African Republic	Ecuador	Guam	Jordan
Belgium	Chad	Egypt, Arab Rep.	Guatemala	Kazakhstan
Belize	Channel Islands	El Salvador	Guinea	Kenya

Countries (K-Z)					
Kiribati	Maldives	New Zealand	Rwanda	St. Lucia	Uganda
North Korea	Mali	Nicaragua	Samoa	St. Martin (French part)	Ukraine
Korea, Rep.	Malta	Niger	San Marino	St. Vincent and the Grenadines	United Arab Emirates
Kosovo	Marshall Islands	Nigeria	Sao Tome and Principe	Sudan	United Kingdom
Kuwait	Mauritania	Northern Mariana Isla	Saudi Arabia	Suriname	United States
Kyrgyz Republic	Mauritius	Norway	Senegal	Swaziland	Uruguay
Lao PDR	Mexico	Oman	Serbia	Sweden	Uzbekistan
Latvia	Micronesia, Fed. Sts.	Pakistan	Seychelles	Switzerland	Vanuatu
Lebanon	Moldova	Palau	Sierra Leone	Tajikistan	Venezuela, RB
Lesotho	Monaco	Panama	Singapore	Tanzania	Vietnam
Liberia	Mongolia	Papua New Guinea	Sint Maarten (Dutch part)	Thailand	Virgin Islands (U.S.)
Libya	Montenegro	Paraguay	Slovak Republic	Timor-Leste	West Bank and Gaza
Liechtenstein	Morocco	Peru	Slovenia	Togo	Yemen, Rep.
Lithuania	Mozambique	Philippines	Solomon Islands	Tonga	Zambia
Luxembourg	Myanmar	Poland	Somalia	Trinidad and Tobago	Zimbabwe
Macao SAR, China	Namibia	Portugal	South Africa	Tunisia	
Macedonia, FYR	Nauru	Puerto Rico	South Sudan	Turkey	
Madagascar	Nepal	Qatar	Spain	Turkmenistan	
Malawi	Netherlands	Romania	Sri Lanka	Turks and Caicos Islands	
Malaysia	New Caledonia	Russian Federation	St. Kitts and Nevis	Tuvalu	